Title
Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review

Permalink
https://escholarship.org/uc/item/4rz0b02v

Authors
Hussain, Mustafa I
Reynolds, Tera L
Zheng, Kai

Publication Date
2019-06-17

DOI
10.1093/jamia/ocz095

Data Availability
The data associated with this publication are in the supplemental files.

Peer reviewed
Review

Medication safety alert fatigue may be reduced via interaction design and clinical role tailoring: a systematic review

Mustafa I Hussain, Tera L Reynolds, and Kai Zheng

Department of Informatics, University of California, Irvine, Irvine, California, USA

Corresponding Author: Mustafa I Hussain, MS, Department of Informatics, University of California, Irvine, 5019 Donald Bren Hall, Irvine, CA 92697, USA (mihussai@uci.edu)

Received 6 March 2019; Revised 14 May 2019; Editorial Decision 15 May 2019; Accepted 19 May 2019

ABSTRACT

Objective: Alert fatigue limits the effectiveness of medication safety alerts, a type of computerized clinical decision support (CDS). Researchers have suggested alternative interactive designs, as well as tailoring alerts to clinical roles. As examples, alerts may be tiered to convey risk, and certain alerts may be sent to pharmacists. We aimed to evaluate which variants elicit less alert fatigue.

Materials and Methods: We searched for articles published between 2007 and 2017 using the PubMed, Embase, CINAHL, and Cochrane databases. We included articles documenting peer-reviewed empirical research that described the interactive design of a CDS system, to which clinical role it was presented, and how often prescribers accepted the resultant advice. Next, we compared the acceptance rates of conventional CDS—presenting prescribers with interruptive modal dialogs (ie, “pop-ups”)—with alternative designs, such as role-tailored alerts.

Results: Of 1011 articles returned by the search, we included 39. We found different methods for measuring acceptance rates; these produced incomparable results. The most common type of CDS—in which modals interrupted prescribers—was accepted the least often. Tiering by risk, providing shortcuts for common corrections, requiring a reason to override, and tailoring CDS to match the roles of pharmacists and prescribers were the most common alternatives. Only 1 alternative appeared to increase prescriber acceptance: role tailoring. Possible reasons include the importance of etiquette in delivering advice, the cognitive benefits of delegation, and the difficulties of computing “relevance.”

Conclusions: Alert fatigue may be mitigated by redesigning the interactive behavior of CDS and tailoring CDS to clinical roles. Further research is needed to develop alternative designs, and to standardize measurement methods to enable meta-analyses.

Keywords: alert fatigue, decision support systems, clinical, medical order entry systems, electronic prescribing, decision support techniques

INTRODUCTION

According to the most recent U.S. government reports, 1 in every 20 deaths in the United States has been attributable to an adverse drug event (ADE). Many ADEs result from erroneous prescriptions. By the most conservative estimates, 1 in every 50 prescriptions is inappropriate.

Clinical decision support (CDS) is intended to reduce prescription error by providing prescribers with automated guidance during computerized order entry. Some have held high hopes for CDS, believing that it would significantly reduce prescription errors. The reality has proved more complex—CDS can create new patient safety risks. For example, in some instances, “hard stops”
have prevented patients from receiving potentially life-saving treatment in time. The information technology infrastructures that organizations must install to integrate CDS into the medication-ordering process—often accompanied by changes in workflow and communication patterns—can disrupt work during “roll-out,” as well as in long-term use. These disruptions can increase instances of ADEs, which can, in turn, increase patient mortality.

CDS can also fail to improve patient safety due to alert fatigue. Alert fatigue occurs when a high number of irrelevant alerts leads users to habitually override them. It is a term derived from alarm fatigue, which psychologists and human factors researchers have used when studying high false alarm rates in fields such as aviation and nuclear power plant operation.

Alert fatigue was once referred to as the “cry-wolf effect” because, much like Aesop’s fable, it describes a situation in which people stop responding to false alarms. Severe consequences can result from alarm and alert fatigue conditions. For example, a 1997 plane crash was attributed to alarm fatigue—the control tower operators had disabled a minimum safe altitude alarm due to its frequent false alarms. Similarly, the patient safety goals of CDS can be compromised by alert fatigue.

Some researchers have focused on increasing alert sensitivity and specificity by modifying CDS ruleset. The results have been mixed. It is often difficult to justify disabling alerts due to safety concerns or pressure from patient safety groups (eg, Leapfrog). Psychologists and human factors researchers have developed strategies to reduce alarm fatigue via interaction design—the design of the way the “dialogue” unfolds between the human user and the computer. Some have applied these strategies in CDS, with promising results. For example, tiered alarms indicate the likelihood or severity of an adverse event, and they seem to have been well received in CDS. As another example, “patient” alarms—those that avoid distracting airplane pilots when they are busy—may be accepted more often than “impatient” alarms. Similarly, in CDS, researchers have implemented alerts that avoided requiring attention at a particular time—again, with some success.

To address whether the interactive design of CDS affects clinical alert fatigue, in the aggregate, we conducted a systematic Preferred Reporting Items for Systematic Reviews and Meta-Analyses review. Existing systematic reviews have tended to focus on prescriber performance and patient outcomes rather than alert acceptance. We only included articles that documented acceptance rates, as defined in the Introduction, or rate of override appropriateness. We identified articles published between 2007 and 2017. Two of the authors (T.L.R., M.I.H.) screened the search results, extracting relevant details (interactive features, clinical roles, acceptance rates, and methods) from included studies for analysis. T.L.R. and M.I.H. met often to ensure consistency.

Eligibility criteria
We included peer-reviewed, English-language articles reporting empirical studies about CDS for medication safety. We included articles that documented acceptance rates, as defined in the Introduction, or enough information to calculate an acceptance rate. We screened more than 1 article documenting the same CDS and setting, we retained the most thorough version.

While screening, we used the following additional criteria. First, as our goal was to understand how prescribers acting of their own free will responded to different interventions, we excluded “hard stops,” which impose heavy time penalties to override, and which therefore materially restrict the prescriber’s range of action. Readers interested in an analysis of hard stops should refer to a 2018 systematic review by Powers et al. Second, we excluded articles that did not describe the interactive design in enough detail to produce a description. Third, we excluded articles in which researchers made global changes to an alerting system, but only reported acceptance rates for those alerts intended to convey the most urgency, for certain drug categories, or for a selected subset of users exposed to the alert; some of these authors may have chosen to report only the most palatable results. If, on the other hand, the researchers set out to improve acceptance of a certain type of alert, like antibiotic stewardship or renal dosing, then reporting the acceptance rate for only those alerts was considered appropriate for our analysis.

Data extraction process
For included articles, we extracted interactive features, the clinical role that received CDS, measurement methods, acceptance rates, and rates of override appropriateness. For articles documenting time series trials of incremental changes to the CDS ruleset, we extracted the last recorded result. If an article reported more than 1 intervention—for example, if the authors compared plain modal dialogs with dialogs that provided additional context—we extracted results from each intervention separately. When an acceptance rate was not directly given, we used the equations provided by McCoy et al to derive an acceptance rate.

Many CDS alerts have been presented as modal dialogs (also known as pop-ups), and some of these have provided a button that indicates “acknowledgment,” which takes no action. Some researchers have considered a clicks of this button to count as “acceptance”—but this, too, may rely on an incorrect assumption. Under alert fatigue, modal dialogs become obstacles, and “acknowledgment” buttons become the work-around.

Additionally, in this review, we paid attention to the clinical role of the recipient of the automated guidance, eg, a prescriber or a pharmacist. Other authors have identified that delivering the right guidance to the right recipient is crucial to the acceptance of the alert.

MATERIALS AND METHODS
We conducted a systematic review according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses model. We started by searching the PubMed, Embase, CINAHL, and Cochrane literature databases. The search terms we used are shown in Table 1. We identified articles published between 2007 and 2017. Two of the authors (T.L.R., M.I.H.) screened the search results, extracting relevant details (interactive features, clinical roles, acceptance rates, and methods) from included studies for analysis.

Defining acceptance
We defined acceptance—our main outcome—as a change to a prescription based on computerized advice. This definition excluded “intention to monitor” and “acknowledgment”—explanations of these concepts follow.

Some CDS alerts have allowed prescribers to select “intention to monitor” as an override justification, and some researchers have counted this justification-selection as evidence of “acceptance.” However, Slight et al. found evidence of monitoring in only 36% of instances in which the prescriber indicated an intention to monitor.
The same 2 authors (T.L.R., M.I.H.) split this data extraction workload evenly and checked one another's work. Doubts on inclusion were handled by interpreting the inclusion criteria in-person to achieve consensus.

Data analysis and synthesis of results

We coded features and measurement methods as short descriptions (eg, “tiered modal dialog presented to the prescriber,” “counted dialog button-clicks”). We sorted these descriptions into categories as commonalities emerged.

We also paid attention to the methods used to construct acceptance rates. In this article, we refer to the 2 main methods as in-dialog action analysis and event analysis.

In-dialog action analysis is only applicable when the CDS intervention takes the form of a dialog that features a button that the prescriber can click to modify or discard their order (eg, “Discard Warfarin Order”). Researchers count the number of times the “acceptance” button was clicked, and divide that count by the total number of dialogs that appeared.

Event analysis may be applied to any form of CDS, including dialogs. When conducting an event analysis, researchers search the patient chart for evidence that the prescriber accepted advice, in addition to any changes that prescribers may have made by clicking buttons inside CDS dialogs. For example, a prescriber might dismiss a modal dialog warning against a warfarin order, and then reduce the dose later. Or, a pharmacist might receive an alert from a CDS system, modal dialog warning against a warfarin order, and then reduce the dose later. Or, a pharmacist might receive an alert from a CDS system, and counsel the prescriber by phone—in which case the researchers must check to see if the prescriber made a change to the chart.

We plotted the frequencies of measurement methods by publication year to examine their popularity over time. For those studies that used more than 1 measurement method, we compared the results of those measurement methods. We also plotted the frequencies of interactive and role-tailoring features reported over time, to identify trends.

Next, we used a t test to compare acceptance rates between CDS systems by interactive design and clinical role-tailoring. In addition, we constructed a plot to holistically examine prescribers’ acceptance rates by feature.

RESULTS

Study selection

As shown in Figure 1, we initially identified 2699 records by querying the literature databases. After removing duplicates, screening titles, and abstracts, and examining full-texts to determine eligibility, we determined that 39 articles met our inclusion criteria. Extracting results from these articles yielded 42 different interventions, since there were 3 articles that reported 2 interventions each.

Study characteristics

The study characteristics are shown in Figure 2. Twenty-four (61%) of the 39 included articles reported studies conducted in the United States, and 3 (8%) reported studies from Taiwan. There were 2 studies from Switzerland, 2 from the Netherlands, and 1 from each of the following: the United Kingdom, China, Canada, and Belgium. Nine of the 24 (38%) studies conducted in the United States were conducted in Harvard-affiliated institutions.

Seventeen (44%) studies were conducted in inpatient settings only, 12 (31%) were conducted in outpatient settings only, 6 (15%) studied both inpatient and outpatient settings, and the remaining 4 (10%) studies were conducted in the emergency department, in the emergency department and outpatient settings, or in an unspecified setting. Twenty-five of the 39 (64%) included articles studied academic healthcare settings, 12 (31%) studied nonacademic settings, and the remaining 2 (5%) studied both settings.

Twenty-five of the 39 (64%) included articles documented an electronic health record (EHR)–integrated CDS, 9 (23%) documented a standalone CDS, and 5 (13%) did not specify whether the CDS was integrated into an EHR. Three of the 39 included articles (8%) reported more than 1 intervention; each intervention was treated as a separate study.

Twenty-four (60%) of the articles solely studied physician behavior. Ten (26%) studied both physician and nurse practitioner behavior and 5 (13%) did not specify the clinical roles that were studied.

Trends in measuring acceptance

As mentioned in the Materials and Methods, we analyzed the methods that researchers used to construct acceptance rates. The number of studies that conducted in-dialog action analyses (n = 23) was approximately equal to the number of that conducted event analyses (n = 22). Eight of the studies in our analysis—all between 2012 and 2017—conducted a review of acceptability, either of the CDS alerts or of overriding behavior, using the method described by Weingart et al35 in 2003.

Three articles contained measurements of prescriber acceptance using both in-dialog action analysis and event analysis. Woods et al36 arrived at an acceptance rate of 26% using in-dialog action analysis.
analysis, and an acceptance rate of 41% using event analysis. Slight et al. arrived at acceptance rates of 40% and 66%, using in-dialog action analysis and using event analysis, respectively; McCoy et al. arrived at acceptance rates of 18% and 47%, respectively. Event analyses generally yielded acceptance rates twice as high (194%) as in-dialog analyses.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis flow diagram.

Figure 2. Study characteristics. ED: emergency department; EHR: electronic health record.
Trends in CDS interventions

Features present in 4 or more included studies are plotted cumulatively, over time, in Figure 3. Three of the most common interactive features—tiering alerts, providing shortcuts for common corrective actions, and requiring a reason to override—are described and illustrated in Table 2.

The most commonly reported type of CDS—which comprised 83% of results—interrupted prescribers with modal dialogs. The most common variants were tiered to convey levels of risk, provided shortcuts for common corrections, or required a reason to override.

We also found advisories that were not automatically issued using computerized systems. These included fax or mail alerts, and...
In this systematic review, we found that interrupting prescribers with modal dialogs have become the least accepted—yet the most prevalent—design. In this section, we analyze possible reasons to account for this observation. Afterward, we discuss some methodological dilemmas faced in CDS research. Some of these have been a matter of methodological inconsistency—and they presented a practical barrier to meta-analysis. Finally, we conclude this section with our recommendations to improve the quality of CDS design and research.

**DISCUSSION**

In this systematic review, we found that interrupting prescribers with modal dialogs have become the least accepted—yet the most prevalent—design. In this section, we analyze possible reasons to account for this observation. Afterward, we discuss some methodological dilemmas faced in CDS research. Some of these have been a matter of methodological inconsistency—and they presented a practical barrier to meta-analysis. Finally, we conclude this section with our recommendations to improve the quality of CDS design and research.

**Reasons why prescriber-interruptive modals seem to elicit alert fatigue**

When we compared prescriber-interruptive modal dialogs with alternatives, we found evidence favoring the alternatives—in particular, those that tailored CDS to the roles of pharmacists and physicians. We believe there are 3 explanations for this finding. The first concerns etiquette—“proper” etiquette often makes advice easier to receive. The second concerns the division of expert labor between prescribers and pharmacists. The third concerns relevance, which comes naturally to humans, but which remains difficult to compute.

**Etiquette.** In the Introduction, we mentioned that psychologists and human factors researchers tend to endorse presenting guidance “politely”—even in emergencies. Prescribers might have accepted pharmacists’ advice so readily because those pharmacists produced behavioral patterns culturally recognized as “polite.” In some cases, it is appropriate to carefully design and program computers to produce similar “behavior” to solicit the user’s reciprocity. Some of the modal dialogs that we saw, which featured large, capitalized red text, and which required several clicks and keystrokes to dismiss, might have been seen as patronizing, rather than polite. In our review, attempts to imitate “politeness” in CDS were rare to find.

**Division of expert labor.** It has been common practice for prescribers to consult pharmacists about the appropriateness of particular medications for patient cases. Presenting certain medication-related CDS to the pharmacist—such as those concerning antibiotic targeting and renal dosing—therefore may support (rather than disrupt) an established clinical practice. This division of expert labor might have a hidden advantage: Sheltering prescribers from most of the details of pharmacy review might allow prescribers to focus more of their attention on the key details of clinical cases, so that they may think more clearly. We understand this may not be feasible in certain cases until regulatory barriers are changed.

**Relevance.** Prescribers may have found pharmacist-mediated CDS alerts highly acceptable because pharmacists filtered out irrelevant advice. Whether computers might, someday, handle relevance and context as capably as humans has been a matter of debate.
Indeed, CDS does not seem to perform at the same level of precision and relevance as the humans they advise.\textsuperscript{51} The prevalence of prescriber-interruptive modal dialogs in the literature might be due to overly narrow definitions of “decision support” by certain patient advocacy groups\textsuperscript{15} or it may be due to actual prevalence in clinics. Additionally, EHR homogenization\textsuperscript{12} may have determined which types of decision support have been convenient for clinical institutions to implement, and which have been expensive, at scale.

CDS homogeneity presented 1 of several barriers to meta-analysis. The other barriers were primarily due to methodological inconsistencies in the literature. Next, we discuss methodological issues.

Mediation analysis may address methodological dilemmas

As mentioned in the Results, we found that researchers had been using 2 main ways to measure how often a prescriber accepted computer-generated advice: in-dialog action analysis and event analysis. Some studies explicitly conducted comparative analyses of the validity of the 2 methods.\textsuperscript{32,36,37} As previously mentioned, when using in-dialog action analysis, the researchers dichotomize the actions taken inside a modal dialog: The prescriber either accepts the alert (eg, by clicking “Discard Order”) or overrides it (eg, by clicking “Proceed Anyway”). We note 3 problems with this method’s validity. First, those clicks provide a rather partial story of the order—for example, they do not account for possible corrections that the prescriber may take after responding to the alert. This is related to the second problem: Applying in-dialog action analysis to modals that feature action shortcuts may artificially inflate acceptance rates with respect to other modal dialogs, because more actions that would otherwise take place outside the dialog would instead take place inside the dialog. Third, in-dialog action analysis cannot be used with CDS interventions that do not offer decision-buttons to prescribers—these interventions must be studied with event analysis.

When using event analysis, the researcher additionally searches for corrective actions that the prescriber made after dismissing any alert, including a modal dialog. For example, the prescriber may change a dose, or switch to a narrow-spectrum antibiotic, after dismissing a modal dialog. These adjustments are taken as evidence of acceptance. This main problem with this method’s validity is that there is no way to know whether the prescriber would have taken the same action if the intervention had not been delivered.

One might expect that in-dialog action analysis errs on the side of specificity (it systematically fails to recognize corrective actions), while event analysis errs on the side of sensitivity (it may misattribute some corrections to interventions). The evidence we gathered from 3 studies that compared these 2 methods\textsuperscript{32,36,37} suggest that this intuition is correct. These methods are biased, in a traditional sense: they produce results that predictably depart from the results that one would expect from the most accurate instrument imaginable.

Despite their limitations, we believe these methods to be valuable, as they seem to be the most cost-effective ways to capture data for CDS acceptance. However, we must caution that these methods produce results that are noncomparable. We suggest using event analysis, to enable rigorous comparisons between modal and modeless forms of CDS.

 Appropriateness panel reviews\textsuperscript{32,35} were rare to see. We imagine these reviews to be particularly costly. Indeed, half of the included articles that reported an appropriateness review were from well-resourced academic institutions.

In fact, the scarcity of information that was useful in our review was surprising given the quantity of available CDS literature. Nearly 9 in 10 of the articles we excluded did not report prescriber acceptance, an important mediating variable between the technological intervention and patient outcomes that seems to have been assumed. Earlier, we described a homogeneity of CDS interventions in the reported literature—specifically, prescriber-interruptive modal dialogs comprised 5 in 6 included results. This also constrained the analyses that could be conducted with adequate statistical power. Finally, a roughly 50-50 split between 2 incomparable measurement methods precluded meta-analysis.

This review revealed several issues in the literature. Future work is needed to develop standardized, low-cost, informative measures for determining acceptance for CDS, and for relating CDS acceptance to patient outcomes. Next, we present some feasible recommendations for improving the quality of the CDS literature.

Recommendations for future work

Given the preceding discussion, we propose the following 3 recommendations for future CDS research:

First, we recommend that researchers consider alternatives to prescriber-interruptive modal dialogs, since there is evidence that the latter suffers from relatively lower acceptance. Role-based tailoring appeared to improve acceptance rates, and further work is needed in this area. Ideally, those who will receive the alerts should be involved in role-tailoring decisions. Alternatives to modal dialogs should also be explored.

Second, recommend measuring acceptance rates using event analysis, rather than in-dialog action analysis. Because event analysis is more widely applicable, using it will enable meta-analyses that accommodate varied CDS interventions.

Last, we recommend reporting both acceptance rates and patient outcomes. Much of the literature that we saw in our review reported one or the other; few reported both. This has made it difficult to analyze patient outcomes as a function of CDS design and role-tailoring, mediated\textsuperscript{53} by acceptance.

CONCLUSION

Alert fatigue remains a persistent challenge in CDS. Among prescriber-interruptive modal dialogs, acceptance rates have been highly variable. In our analysis, prescribers accepted alternative interventions more often—especially those which tailored CDS to the areas of expertise associated with clinical roles. Although there are plausible reasons why some alternative CDS interventions would improve acceptance, contemporary literature has not supported detailed analyses. We recommend that future studies pay more attention to alternative designs, measure acceptance using event analysis, and report patient outcomes as well as acceptance rates.

FUNDING

The project described was supported in part by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through Grant (UL1 TR001414). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. The project was also supported in part by
Academic Senate Council on Research, Computing and Libraries (CORCL), and by the U.S. Department of Education (DoE), Graduate Assistance in Areas of National Need (GAANN).

AUTHOR CONTRIBUTIONS
Mustafa I. Hussain provided substantial contributions to conception, design, data acquisition, analysis, interpretation, and drafting. Tera L. Reynolds provided substantial contributions to the acquisition and analysis of data for the work, as well as critical revisions for important intellectual content. Kai Zheng provided substantial contributions to the conception of the work, and the interpretation of data for the work, as well as critical revisions for important intellectual content. All authors provided final approval of the version to be published, and have agreed to be accountable for all aspects of the work.

SUPPLEMENTARY MATERIAL
Supplementary material is available at Journal of the American Medical Informatics Association online.

ACKNOWLEDGMENTS
The authors would like to thank the Editors and Reviewers for their careful consideration of this work.

Conflict of interest statement. None declared.

REFERENCES